

MA-250 Probability and Statistics

Nazar Khan

PUCIT

Lecture 28

Box Models

- Today's lecture covers Chapters 16 – 18 from the Statistics book by Freedman.

Box Models

- Chance variability can be analysed through a box model.
- A complicated chance process for generating a number can be modeled by drawing from a box.
 - The sum of draws is a key ingredient.
- Example: Sum of N rolls of a fair die corresponds to
 - drawing N times with replacement
 - from the box $[1,2,3,4,5,6]$
 - and taking the sum

Box Models

- Example: Sum of N rolls of a fair die corresponds to
 - drawing N times
 - from the box $[1,2,3,4,5,6]$
 - and taking the sum
- Let random variable X_i represent the number obtained on the i -th draw.
- Then sum $Y = X_1 + \dots + X_N$ where the X_i 's are independent and identically distributed (i.i.d) with mean 3.5 and variance 2.917.
- By the CLT: the sum of N draws will have mean $3.5 * N$ and variance $2.917 * N$ and it will have a normal distribution. (**What was the distribution of each X_i ?**)

Box Models

- For gambling problems in which the same bet is made several times, a box model can be set up
 - Contents of the box show the amounts that can be won (+) or lost (-).
 - Chance of drawing an item from the box equals the chance of winning that amount on a single play.
 - Number of draws = number of plays.
 - Net gain after playing N times = sum of N draws with replacement.
- Since draws are independent and the same experiment is repeated N times, the corresponding set of random variables will be i.i.d.
- So for large N , CLT will apply on the net gain (= sum of draws).

Box Models

Example

- Sum of 25 draws from the box $[0,2,3,4,6]$.
- X_i =value on a i -th draw
 - $E(X_i)=3$
 - $SD(X_i)=2$ ($Var(X_i)=4$).
- $Y=X_1+...+X_n$
- By CLT:
 - $E(Y)=3*25=75$,
 - $Var(Y)=4*25=100$, $SD(Y)=2\sqrt{25}=10$
 - $Y \sim N(75, 10^2)$
- This means that roughly 68% of the time, the sum of 25 draws from this box will be in the range 75 ± 10 .
- In other words, $P(65 \leq Y \leq 85) \approx 0.68$.

Box Models

1. One hundred draws will be made at random with replacement from the box

1	1	2	2	2	4
---	---	---	---	---	---

.

- (a) The smallest the sum can be is _____, the largest is _____.
- (b) The sum of the draws will be around _____, give or take _____ or so.
- (c) The chance that the sum will be bigger than 250 is almost _____%.

2. One hundred draws will be made at random with replacement from the box

1	3	3	9
---	---	---	---

.

- (a) How large can the sum be? How small?
- (b) How likely is the sum to be in the range from 370 to 430?

6. One hundred draws are going to be made at random with replacement from the box

0	2	3	4	6
---	---	---	---	---

. True or false and explain.

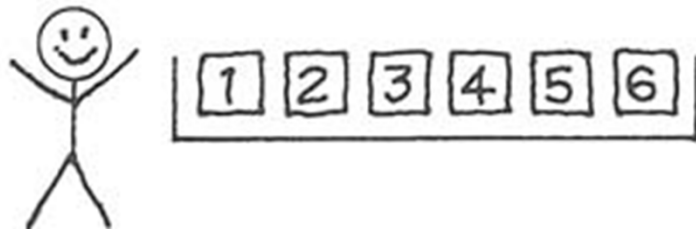
- (a) The expected value for the sum of the draws is 300.
- (b) The expected value for the sum of the draws is 300, give or take 20 or so.
- (c) The sum of the draws will be 300.
- (d) The sum of the draws will be around 300, give or take 20 or so.

Box Models

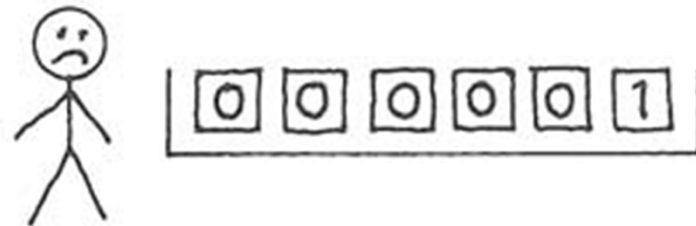
Example 4. A die is rolled 60 times.

- (a) The total number of spots should be around _____, give or take _____ or so.
- (b) The number of 6's should be around _____, give or take _____ or so.

For adding up the draws,
the box is



For counting 6's,
the box is



Box Models

- A coin is tossed 100 times. Estimate the chance of getting between 40 and 60 heads?

Box Models

1. A coin is tossed 16 times.

(a) The number of heads is like the sum of 16 draws made at random with replacement from one of the following boxes. Which one and why?

(i)

head	tail
------	------

 (ii)

0	1
---	---

 (iii)

0	1	1
---	---	---

(b) The number of heads will be around _____, give or take _____ or so.

2. One hundred draws are made at random with replacement from

1	2	3	4	5
---	---	---	---	---

. What is the chance of getting between 8 and 32 tickets marked “5”?

3. According to the simplest genetic model, the sex of a child is determined at random, as if by drawing a ticket at random from the box

male	female
------	--------

What is the chance that of the next 2,500 births (not counting twins or other multiple births), more than 1,275 will be females?

Box Models

13. A letter is drawn 1,000 times, at random, from the word A R A B I A. There are two offers.

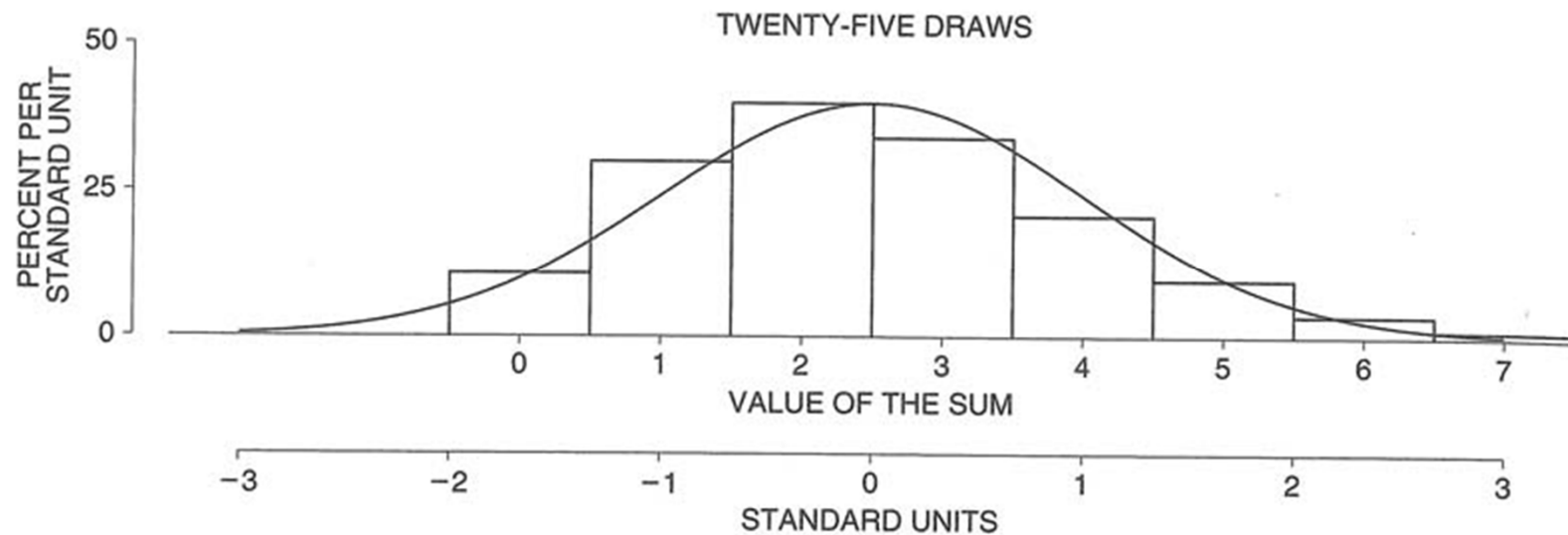
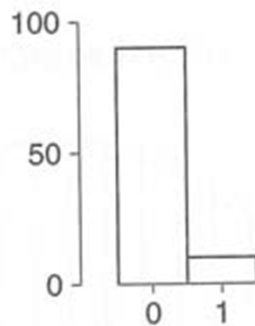
- (A) You win a dollar if the number of A's among the draws is 10 or more above the expected number.
- (B) You win a dollar if the number of B's among the draws is 10 or more above the expected number.

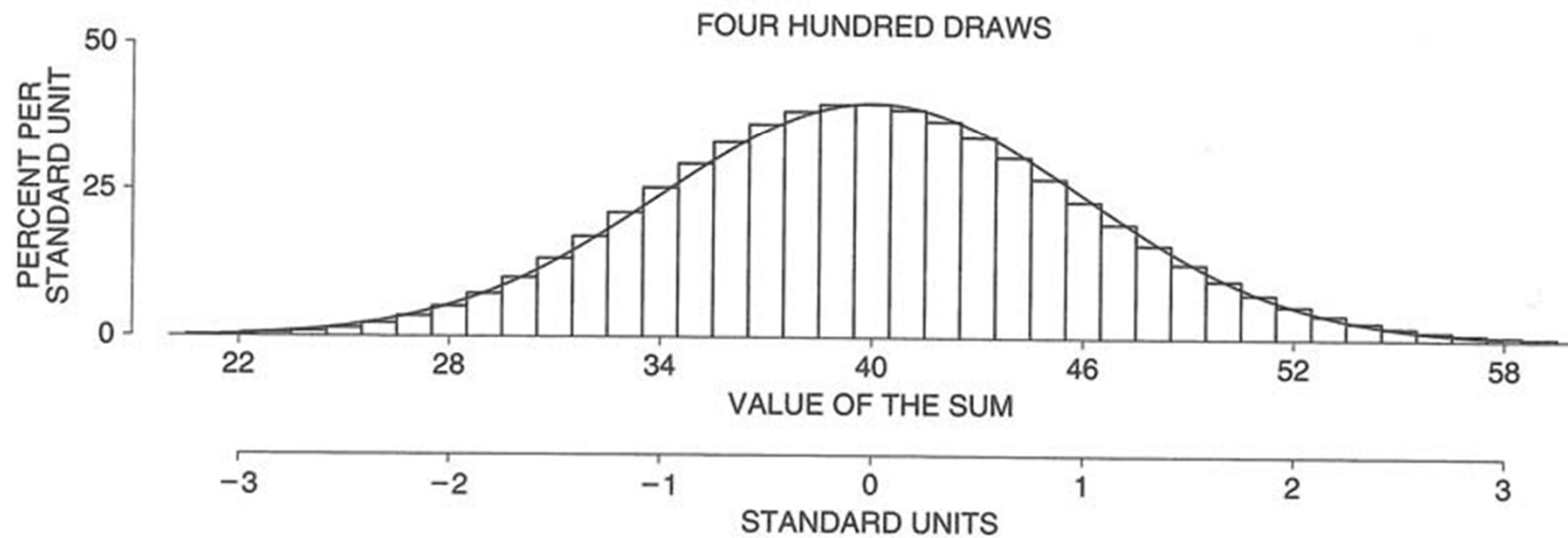
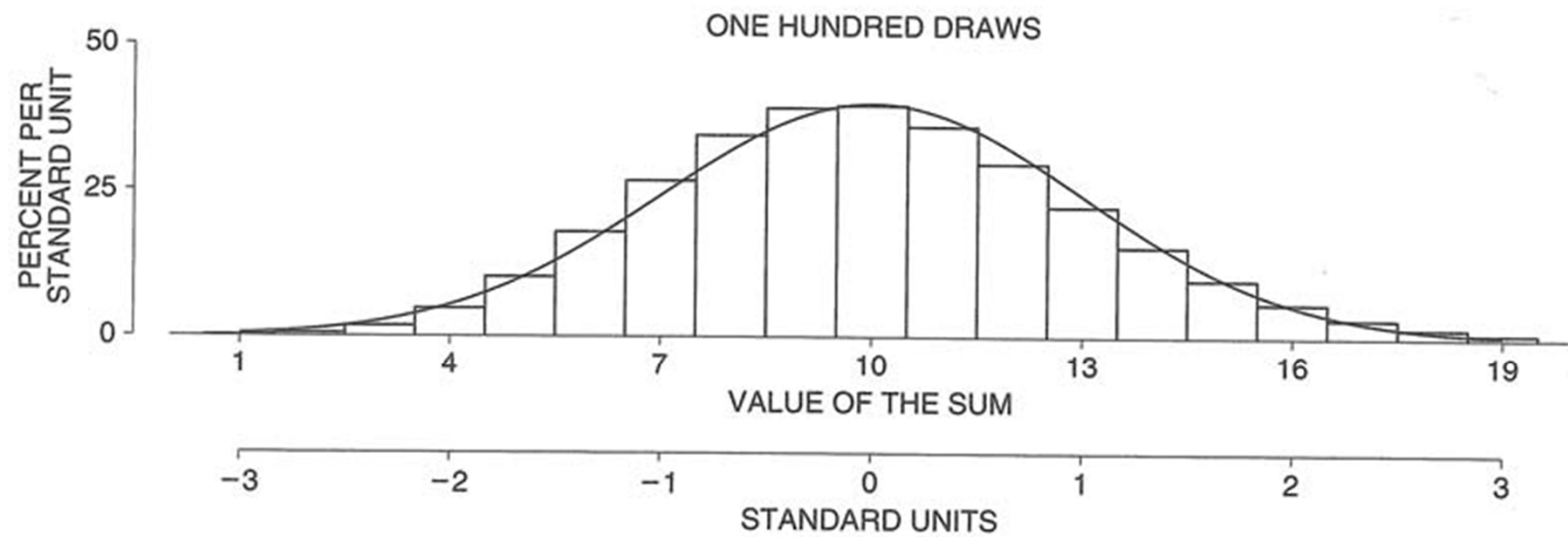
Choose one option and explain.

- (i) A gives a better chance of winning than B.
- (ii) A and B give the same chance of winning.
- (iii) B gives better chance of winning than A.
- (iv) There is not enough information to decide.

Demonstration of the CLT

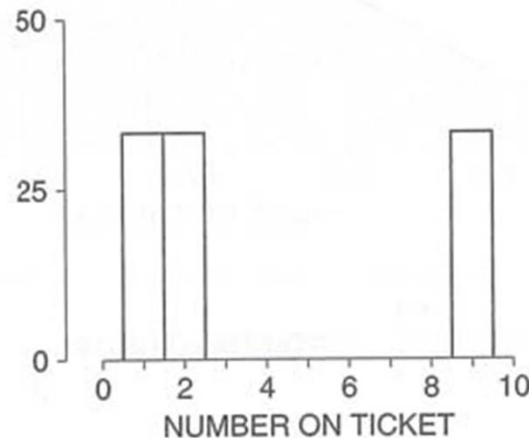
Histogram for the lopsided box $\left[\begin{array}{c} 9 \\ 0 \end{array} \right]$'s $\left[\begin{array}{c} 1 \\ 1 \end{array} \right]$.

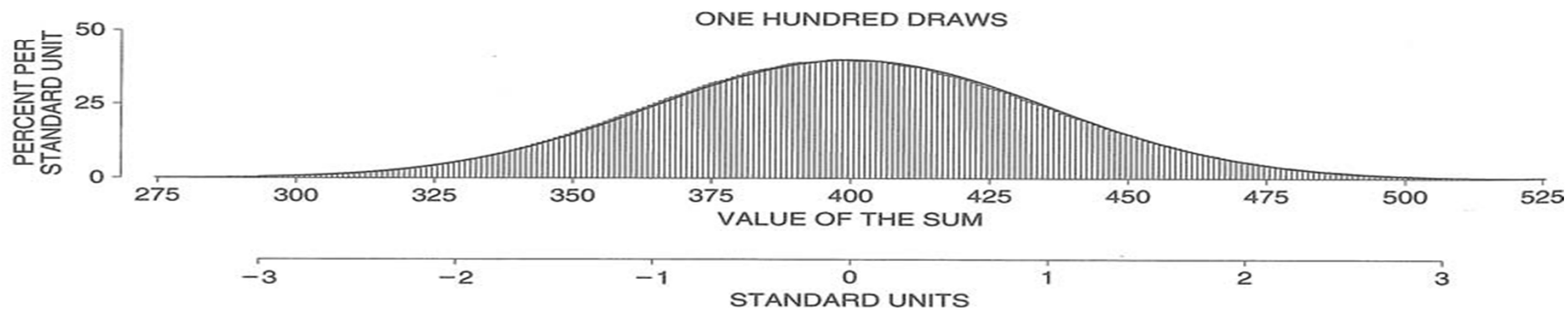
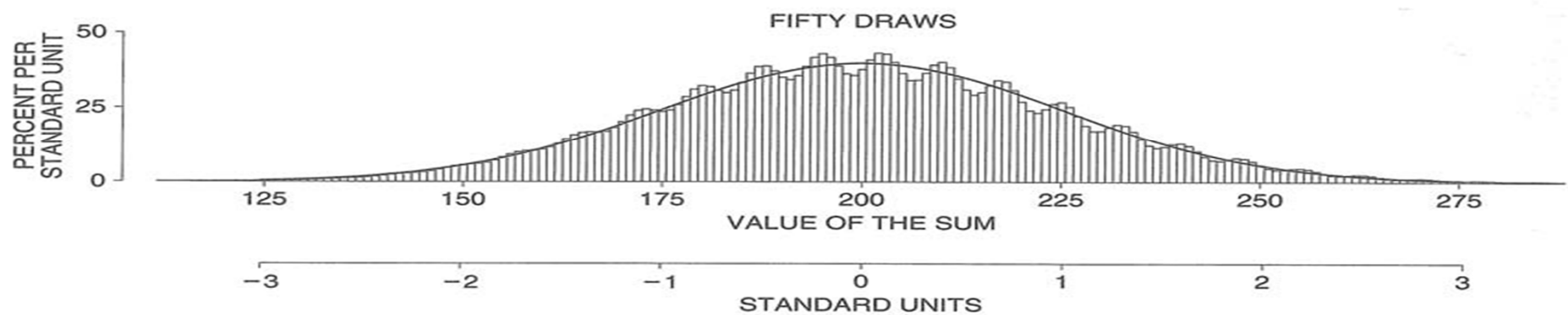
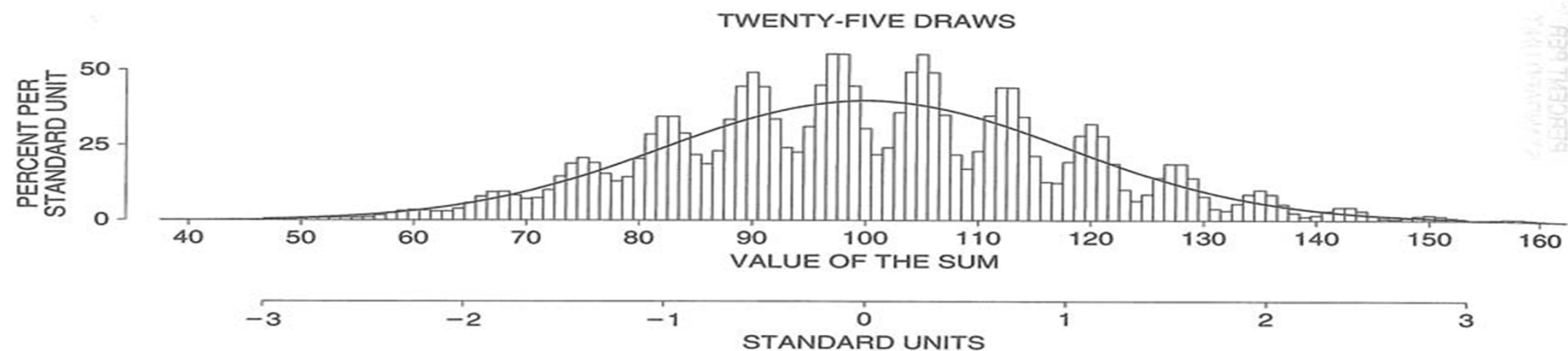




Demonstration of the CLT

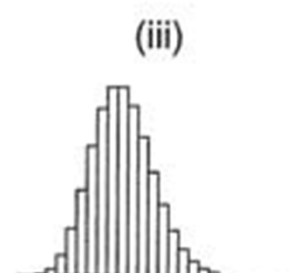
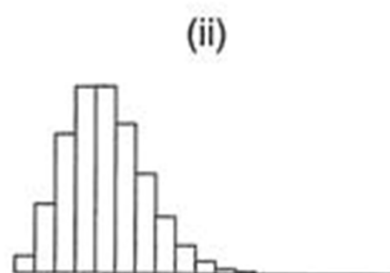
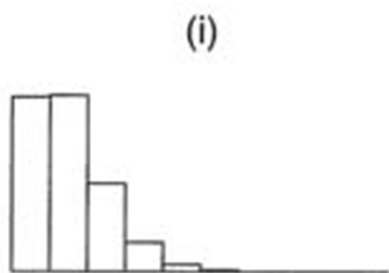
- Drawing with replacement from the box $[1,2,9]$.
- Histogram looks nothing like the normal curve.





2. A biased coin has one chance in ten of landing heads. It is tossed 400 times. Estimate the chance of getting exactly 40 heads.

6. Shown below are probability histograms for the sum of 100, 400, and 900 draws from the box $\begin{array}{|c|} \hline 99 \text{ } 0 \\ \hline \end{array}$'s $\begin{array}{|c|} \hline 1 \\ \hline \end{array}$. Which histogram is which?



2. Four hundred draws will be made at random with replacement from the box

1	3	5	7
---	---	---	---

.

- (a) Estimate the chance that the sum of the draws will be more than 1,500.
- (b) Estimate the chance that there will be fewer than 90

3

's.

3. Ten draws are going to be made at random with replacement from the box

0	1	2	3
---	---	---	---

. The chance that the sum will be in the interval from 10 to 20 inclusive equals the area under _____ between _____ and _____. Fill in the blanks. For the first one, your options are: the normal curve, the probability histogram for the sum. Explain your answers.

4. A coin is tossed 25 times. Estimate the chance of getting 12 heads and 13 tails.

9. One hundred draws are made at random with replacement from a box with ninety-nine tickets marked "0" and one ticket marked "1." True or false, and explain:

- (a) The sum will be around 1, give or take 1 or so.
- (b) There is about a 68% chance that the sum will be in the range 0 to 2.